CECR Center for Educator Compensation Reform

Using Value-Added Indicators for Measuring School Improvement

Written by Robert H. Meyer, Director of the Value-Added Research Center and Senior Scientist at the Wisconsin Center for Education Research, University of Wisconsin–Madison

Educational outcome indicators are routinely used to measure the performance of schools, programs, and policies. Such indicators will be used, at least in part, to determine the compensation of teachers and principals in the Teacher Incentive Fund projects. This article discusses the weaknesses of the most commonly used educational outcome indicators—average test scores and proficiency rates—and the advantages of value-added indicators for the specific purpose of measuring the productivity of schools as well as classrooms and teachers.[1] Several major conclusions emerge from the analysis.
Attainment indicators, such as average test scores or proficiency rates (even if they are derived from highly valid assessments) provide institutions with the perverse incentive to "cream"—that is, to raise measured performance by educating only those students who tend to have high test scores. The potential for creaming is apt to be particularly strong in environments characterized by selective admissions. However, creaming also could exist in subtler, but no less harmful, forms. For example, schools and programs could create an environment that is relatively unsupportive for potential dropouts, academically disadvantaged students, and special education students, thereby encouraging these students to drop out, transfer to another school, or enroll in a different program. Other potentially negative impacts of attainment indicators include schools aggressively retaining students at given grade levels as well as high-quality teachers and administrators gravitating to schools and programs that predominantly serve high-scoring students.

Moreover, attainment indicators tend to be biased against schools and programs that disproportionately serve academically disadvantaged students. One source of bias is the well-known fact that school productivity is only one of the many determinants of student achievement. Much of the variation in average or median test scores usually can be accounted for by differences across schools in student achievement prior to students entering a school or to the types of students enrolled.

The Value of the Value-Added Approach

Given the substantial problems that exist with attainment indicators as measures of school productivity, what are the feasible alternatives? There is a growing consensus that the most appropriate method for measuring the school as well as the classroom or teacher is the value-added approach. The essence of the value-added approach is that school, classroom or teacher, or program performance is measured using a statistical regression model that includes, to the extent possible, all of the nonschool factors that contribute to growth in student achievement—in particular, prior student achievement and student and family characteristics.[2] The key idea is to statistically isolate the contribution of schools and programs to growth in student achievement at a given grade level from all other sources of

[1] Many of the issues discussed in this article are considered at greater length in Meyer (1996).
[2] Student and family characteristics could be measured directly or indirectly using repeated observations on students (longitudinal data).

student achievement growth.[3] This is particularly important in light of the fact that differences in prior achievement and student and family characteristics account for far more of the variation in student achievement than school-related factors. Failure to account for differences across schools in these characteristics could result in highly contaminated indicators of performance.

Additional Information About the Weakness of Attainment Indicators as Measures of School Productivity

A school-level attainment indicator, such as an average test score or a proficiency rate, is a flawed measure of school performance for the following four basic reasons:

- **Lack of Localized School Performance to the Classroom or Grade Level.** The attainment indicator fails to localize school performance to a specific classroom or grade level—the natural unit of accountability in a traditional school. This lack of localization is, of course, most severe at the highest grade levels. A performance indicator that fails to localize school performance to a specific grade level or classroom is likely to be a relatively weak instrument of public accountability.
- **Out-of-Date Information About School Performance.** The attainment indicator reflects information about school performance that tends to be grossly out-of-date. Consider, for example, the attainment indicator for a group of students tested at the end of 10th grade. The attainment indicator for this group is a reflection of the accumulated learning that occurred in 10th grade during the prior year, in ninth grade—two years earlier, in eighth grade—three years earlier, and so on, all the way to kindergarten and preschool—11 (or more) years earlier. Indeed, a 10th-grade-level indicator could be dominated by information that is five or more years old. One consequence of this situation is that changes over time in attainment indicators could be negatively correlated with actual changes in program performance (Meyer, 1996). The fact that attainment indicators reflect out-of-date and possibly misleading information severely weakens them as instruments of public accountability. To allow educators to react in a timely and responsible fashion, performance indicators must reflect information that is current and accurate.
- **Contamination Due to Student Mobility.** Attainment indicators at the school, district, and state levels tend to be highly contaminated due to student mobility. For example, the typical high school student is likely to attend several different schools over the period spanning kindergarten through 12th grade. For these students, a test score reflects the contributions of more than one and possibly many different schools. The problem of contamination is compounded by the fact that rates of student mobility tend to differ dramatically across schools. Contamination is apt to be especially high in communities that undergo rapid population growth or decline as well as in communities that experience significant changes in their occupational and industrial structure. Contamination due to student mobility is probably a relatively minor problem at the national level, because rates of migration in and out are low compared to rates of mobility within the nation; but, at the district and school levels, it is apt to be quite serious.

[3] Note that value-added indicators focus on the growth in student achievement from one grade to the next for given cohorts of students rather than on the change (or trend) over time in average test scores for students at a given grade level. Value-added indicators are thus based on longitudinal as opposed to cross-sectional student data.

- **Contamination by Factors Other Than School Performance.** The attainment indicator is contaminated by factors other than school performance, in particular, the average level of student achievement prior to entering first grade (average initial achievement) and the average effects of student, family, and community characteristics on student achievement growth from first grade through the grade in which students are tested. In fact, it is quite likely that comparisons across schools of attainment indicators primarily reflect these differences rather than genuine differences in intrinsic school performance. As such, attainment indicators are highly biased against schools that disproportionately serve academically disadvantaged students and communities.

An Example Based on National Data

The practical significance of the previously described analysis is illustrated using data on average mathematics scores from 1973 to 1986 from the National Assessment of Educational Progress (NAEP). As indicated in Panel A of Table 1, NAEP scores for Grade 11 exhibit the by now-familiar pattern of sharp declines from 1973 to 1982 and then partial recovery between 1982 and 1986. The Grade 11 data, by themselves, are fully consistent with the premise that academic reforms in the early and mid-1980s generated substantial gains in academic achievement. In fact, an analysis of the data based on a *gain indicator* (a value-added type indicator) rather than an attainment indicator suggests the opposite conclusion. (Refer to Panel B of Table 1.)

The gain indicator is similar to a true value-added indicator in that it controls for differences among students in prior achievement. It does so in a very simple and intuitive way: Gain is the change in attainment indicators over time (and across grades) for the *same cohort* of students. For example, the gain in test scores for students who were in Grade 11 in 1986 is given by attainment indicator of Grade 11 students in 1986 minus the attainment indicator for Grade 7 students in 1982 (four grades and four years earlier) (that is, 302.0 – 268.6 = 33.4). Unfortunately, the gain indicator, unlike the value-added indicator, does not control for differences in student, family, and neighborhood characteristics that contribute to growth in student achievement. As a result, the gain indicator reflects possible changes over time in the composition of the population as well as changes in school productivity.[4] Nonetheless, it is instructive to compare the gains in achievement experienced by different cohorts.[5] For this illustrative analysis, we assume that average test scores in 1973 are comparable to the unknown 1974 scores.

As indicated in Panel B, the achievement growth of high school students (from Grade 7 to Grade 11) during the 1982 to 1986 period was actually no better than achievement growth during previous periods. In fact, the gain from Grade 7 to Grade 11 was actually slightly lower during the 1982 to 1986 period than in previous periods! The rise in Grade 11 mathematics scores from 1982 to 1986 stems from an earlier increase in achievement growth for this cohort of students rather than from an increase in achievement growth over Grade 7 to Grade 11. In short, these data provide no support for the notion that high school academic reforms generated significant increases in test scores during the mid-1980s.

[4] The gain indicator also cannot be constructed if the tests before (pretests) and after (posttests) differ and have not been placed on the same measuring scale.

[5] NAEP was originally designed to permit this type of analysis. In mathematics, the tests have generally been given every four years at grade levels spaced four years apart.

These data also vividly confirm the general superiority of the gain indicator, relative to level indicators such as the attainment indicator, as a measure of educational productivity.

Table 1. NAEP Mathematics Examination Data

(A) Average Test Scores by Year

Grade	1973	1978	1982	1986
Grade 3	219.1	218.6	219.0	221.7
Grade 7	266.0	264.1	268.6	269.0
Grade 11	304.4	300.4	298.5	302.0

(B) Average Test Score Gain From Year to Year for Each Cohort

Grade	1973–1978	1978–1982	1982–1986
Grades 3–7	45.0	50.0	50.0
Grades 7–11	34.4	34.4	33.4

Source: Dossey, Mullis, Lindquist, and Chambers (1988)

Summary

Attainment indicators such as the average test score or proficiency rate, the most commonly used indicators in American education, are highly suspect as indicators of school and program performance. These indicators suffer from four major deficiencies: (1) They fail to localize performance to the classroom or grade level; (2) they aggregate information on performance that tends to be grossly out-of-date; (3) they are contaminated by student mobility; and (4) they fail to measure the distinct contribution of schools and programs to growth in student achievement as opposed to the contribution due to students, families, and community factors. As a result, they are flawed measures for evaluation purposes and are weak, if not counterproductive, instruments of public accountability.